TITLE: Responsiveness and Interpretability of 2 Measures of Physical Function in Patients With Spondyloarthritis

RUNNING HEAD: Measuring Physical Function in Spondyloarthritis

TOC CATEGORY: Measurement

ARTICLE TYPE: Original Research

AUTHOR BYLINE: Camilla Fongen, Hanne Dagfinrud, Annelie Bilberg, Elisabeth Pedersen, Melissa Woll Johansen, Salima van Weely, Kåre Birger Hagen, Silje Halvorsen Sveaas

AUTHOR INFORMATION:

C. Fongen, PT, MSc, Department of Rheumatology, National Advisory Unit on Rehabilitation in Rheumatology, Diakonhjemmet Hospital, Oslo, Norway.

H. Dagfinrud, PT, Department of Rheumatology, National Advisory Unit on Rehabilitation in Rheumatology, Diakonhjemmet Hospital.

A. Bilberg, PT, PhD, Institute of Neuroscience and Physiology, Section of Health and Rehabilitation, Physiotherapy, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, and Department of Physiotherapy, Sahlgrenska University Hospital, Gothenburg, Sweden.

E. Pedersen, PT, Department of Rheumatology, University Hospital of North Norway, Tromsø, Norway.

M.W. Johansen, PT, Department of Rheumatology, Martina Hansens Hospital, Bærum, Norway.

S. van Weely, PT, PhD, Department of Orthopaedics, Rehabilitation and Physical Therapy, Leiden University Medical Center, Leiden, the Netherlands.

K.B. Hagen, PT, Department of Rheumatology, National Advisory Unit on Rehabilitation in Rheumatology, Diakonhjemmet Hospital.

S.H. Sveaas, PT, PhD, Department of Rheumatology, National Advisory Unit on Rehabilitation in Rheumatology, Diakonhjemmet Hospital, PO Box 23 Vinderen, 0319 Oslo, Norway. Address all correspondence to Dr Sveaas at: s.h.sveaas@medisin.uio.no.

KEYWORDS: Axial Spondyloarthritis, Physical Function Assessments, Responsiveness, Interpretability

ACCEPTED: October 6, 2019

SUBMITTED: April 4, 2019

Background. Maintenance or improvement of physical function is an important treatment target in the management of patients with axial spondyloarthritis (axSpA); measurement tools that can detect changes in physical function are therefore important.

Objectives. The objective of this study was to compare responsiveness and interpretability of the patient reported Bath Ankylosing Spondylitis (AS) Functional Index (BASFI) and the AS performed based Improvement (ASPI) in measuring change in physical function after exercise in patients with axSpA.

Design. This was a sub-study of 58 patients nested within a randomized controlled trial comparing the effect of 12-weeks exercise with usual care.

Methods. Responsiveness and interpretability was assessed according to the Consensus-based Standards for the selection of health status Measurement Instrument (COSMIN). Responsiveness was assessed by testing eight predefined hypotheses for ASPI and BASFI. Interpretability was assessed by (1) using patients' reported change as an anchor ("a little better" = minimal important change [MIC]) and (2) by categorizing patients with a 20% improvement as responders.

Results. For ASPI and BASFI; 5 of 8 (63%) vs. 2 of 8 (25%), of the predefined hypotheses for responsiveness were confirmed. The MIC values for improvement in physical function were 3.7 sec in ASPI and 0.8 points (on a scale from 0-10) for BASFI. In the intervention group**,** 21 of 30 (70%) and 13 of 30 (43%) of the patients were categorized as responders measured with ASPI and BASFI, respectively. There was a tendency towards a floor effect in BASFI, as 8 of 58 (14%) patients scored the lowest value at baseline.

Limitations. This study was limited by its moderate sample size.

Conclusions. Our findings suggest that ASPI is preferable over BASFI when evaluating physical function after exercise interventions in patients with axSpA.

Axial spondyloarthritis (axSpA) is a chronic inflammatory rheumatic disease that mainly affects the sacroiliac joints and spine.[1] The main clinical features include inflammatory back pain, progressive restriction in spinal mobility, joint stiffness, fatigue, and it is also associated with arthritis, enthesitis and extra-articular manifestations.[1] Maintaining or improving physical function is defined as an important treatment goal for this patient group, and

physical therapy with exercise is recommended as a cornerstone in management.[2] Hence, outcome measures that can detect changes in physical function are needed to evaluate the effectiveness of exercise interventions in patients with axSpA.

The Assessment of SpA international Society (ASAS) recommend the use of the patient reported disease-specific Bath Ankylosing Spondylitis (AS) Functional Index (BASFI) to evaluate physical function in patients with axSpA.[3,4] BASFI was launched in 1994, is frequently used, has adequate measurement properties and reflect the patients' perspective.[5] However, the patient group has changed during the last decade; patients are diagnosed at an earlier stage and the introduction of biological medication has revolutionized the treatment and the disease course.[6] Hence, patients' overall physical function has improved since the development of the BASFI. In line with this, it has been shown that the BASFI is less sensitive to detect changes in patients in relatively well patients, and in trials of physical therapy interventions.[5]

To meet the shortcomings of BASFI, a performance-based test derived from BASFI has been developed; the AS Performance-based Improvement (ASPI). ASPI is reported to be easy to administer, well tolerated by patients with varying limitation in physical functioning, and feasible in daily clinical practice. In addition, ASPI has shown adequate measurement properties to evaluate changes in physical function after pharmacological treatment in patients with axSpA,[7-10] but whether ASPI can be used to detect changes after physical therapy intervention has not been investigated. Hence, the aim of the present study was to compare the responsiveness and interpretability of the performance-based measure ASPI and the patient reported measure (BASFI) in measuring change in physical function after an exercise intervention in patients with axSpA.

[H1]METHODS

[H2] Design

This methodological study was a sub-study nested within the Exercise for SpondyloArthritis (ESpA)-study,[11] and includes patients recruited after mid of November 2015 until September 2016. The ESpA-study was an assessor-blinded, multicenter, randomized controlled trial performed in compliance with the Helsinki agreement, and all participants provided their written informed consent before participation. The trial was approved by Regional Committee for medical and health Ethics of South East Norway (2015/86), the Regional Ethical Review Board Gothenburg in Sweden (032-16) and is listed in ClinicalTrials.gov (NCT02356874).

[H2] Participants

The patients were recruited through outpatient rheumatology departments in Norway and Sweden: Diakonhjemmet Hospital, Martina Hansens Hospital, Northern Norway University Hospital and Sahlgrenska University Hospital. Patients with axSpA according to the ASAS classification criteria,[12] age 18 to 70 years, with stable medication for ≥3 months, moderate to high disease activity (BASDAI ≥3.5 or patient global score ≥3.5) and did not participate in structured endurance and strength exercise program during the last 6 months (>1 hour per week) were included. Exclusion criteria were severe co-morbidity involving reduced exercise capacity and/or contraindications for physical activity as per American College of Sports Medicine (ACSM) guideline,[13] not able to participate in weekly exercise sessions and pregnancy.

[H3]*Exercise program*

The intervention was a high intensity exercise program that lasted for 12-weeks, and followed the American College of Sports Medicine (ACSM) exercise recommendations.[13] Two days a week the exercise sessions were supervised. The participants performed high intensity interval exercise on a treadmill or ergometer bike (four minutes at 90-95% of maximal heart rate followed by three minutes of active resting at 70% of maximal heart rate repeated four times). Thereafter, the participants performed 20 minutes of strength exercises for major muscle groups (8-10 repetitions maximum, 2-3 sets). Once a week the participants individually performed a cardiorespiratory exercise session for 40 minutes (described in details in the report of the primary outcome [11]). Participants in the control group received no intervention and were asked to not change physical activity habits.

[H2] Procedure of measurements

All included participants were assessed by a blinded assessor at baseline before randomization, and after the intervention period. The clinical examination followed the same order at each time point. Starting with a laboratory evaluation of C-reactive protein (CRP), measurements of mobility, followed by the performed based test of physical function and at last the cardiorespiratory fitness test. In addition, the patients' completed a standardized set of questionnaires of patient reported physical function, questions about sociodemographic information and disease symptoms.

[H2] Two Measurements of Physical function

[H3]*Performance-based physical function.* Performance-based physical function was tested with ASPI, which is developed to evaluate changes in physical function in patients with AS.[7] ASPI consists of three tests; 1) bending and picking up pens from the floor, 2) putting on socks and 3) getting up from the floor. Patients are instructed to perform all tests as quickly

as possible, though in a safe manner. Outcome of the performance test is the time needed to complete the tasks, measured in seconds.[8] Test 3 (getting up from the floor) was performed tree times and the mean performance time was used. ASPI has been shown to be a reliable, valid and responsive method to evaluate changes in physical function after treatment with TNFi in patients with AS.[7-10]

[H3]*Patient-reported physical function.* Patient-reported physical function was assessed with BASFI,[4] which was developed to measure physical function in patients with AS. BASFI includes 8 questions regarding activities of daily living (putting on sock, bending, reaching, getting up from chair, getting up off the floor, standing, climbing steps, looking over shoulder) and two questions addressing the ability to cope with everyday life (physical demanding activities and full day activity). Each question is answered on an 11-point numeric rating scale (NRS), anchored by easy (0) and impossible (10). The item-scores are summarized and presented as mean BASFI score (0-10, 10=most limited physical function). BASFI has been shown to be a reliable, valid and responsive measure of physical function in patients with axSpA.[5,14] It is recognized that BASFI may be less sensitive to detect changes in patients without reduced spinal mobility and in exercise trials.[5,14]

[H3]*Patient's assessment of change in physical function.* At the end of the intervention period, patients were asked to rate the extent to which their physical function had changed since baseline on a global ranking scale (GRS) with a five point Likert scale; worse, no change, a little better, much better and very much better.

[H2] Other measurements

[H3]*Mobility.* Mobility was assessed with Bath AS Metrology Index (BASMI), which include four measures of spinal mobility (lateral spinal flexion, cervical rotation, lumbar flexion,

tragus-to-wall distance) and one measure of hip mobility (intermalleolar distance).[15] BASMI was developed to quantify the mobility of the axial skeleton in patients with AS. The included measures were collected following the recommendations from ASAS.[3] The formula for BASMI linear was used to compute the total score (0-10, 10=most impairment).[16] The BASMI has been shown to be reliable,[17,18] valid [15,19,20], and the highest responsiveness is shown for the BASMI linear formula.[16]

[H3]*Cardiorespiratory fitness.* Cardiorespiratoty fitness was tested with a maximal walking test on a treadmill, according to the modified Balke protocol.[21] This protocol has previously been used in patients with axSpA and was performed in accordance with previously description.[22,23] The estimated peak oxygen uptake (VO_2peak) was calculated based on the ACSM formula.[24] Estimating VO_2 peak from a maximal test is considered the second most valid test for cardiorespiratory fitness after measurement of VO_2 by ergo- spirometry during a maximal test.[24] Further, a maximal test for cardiorespiratory fitness is known to be responsive for exercise interventions.[25]

[H3]*Disease activity.* Disease activity was assessed with Bath AS Disease Activity Index (BASDAI),[26] including six questions on perceived symptoms (fatigue, back pain, peripheral joint pain, enthesitis related pain, severity and duration of morning stiffness). The questions are answered on an 11-point NRS, anchored by none (0) and very severe (10), except duration of morning stiffness which is anchored by 0 hours (0) and 2 or more hours (10). The scores on question about severity and duration of morning stiffness are averaged and added to the mean of the remaining four questions to give a final score (0-10, 10 = worst disease activity). BASDAI is considered to be a reliable, valid and responsive measure of disease activity in patients with axSpA.[14,27] For background characteristics, disease activity was also

assessed with the composite index AS Disease Activity Score (ASDAS).[28,29] For ASDAS; cut-offs are defined as inactive disease <1.3, low disease activity 1.3 to <2.1, high disease activity 2.1 -3.5, and very high disease activity >3.5.[30]

[H3]*Disease-related symptoms.* Disease-related symptoms were assessed with single questions, as endorsed by ASAS to evaluate the effect of clinical studies in patients with axSpA.[3] Back pain was assessed with the question "How much pain of your spine due to axSpA do you have?" and fatigue was assessed with the question "How would you describe the overall level of fatigue/tiredness you have experienced?", both questions is answered on an 11-point NRS scale, anchored by no pain/fatigue (0) and most severe pain/fatigue (10).[3] The NRS-back-pain and fatigue scores are shown to be reliable, valid and responsive measures in patients with axSpA.[14,31]

[H2] Evaluation of responsiveness

Responsiveness was defined as the ability of the instruments to detect changes over time in physical function[32] and was assessed according to the Consensus-based Standards for the selection of health status Measurement Instruments (COSMIN).[33] The approach used to assess responsiveness was to postulate and test predefined hypotheses formulated by an expert group in analogy to construct longitudinal validity.

We established an expert group including researchers, experienced physiotherapists and a rheumatologist. The group defined eight a priori hypotheses, addressing expected correlations between change scores in physical function and change scores between physical function and other measures such as spinal mobility, cardiorespiratory fitness, disease activity, back pain and fatigue. As correlation coefficients are reported to be lower in responsiveness studies than in studies of construct validity,[33] the correlations between

change scores in measurements of similar and related constructs were hypothesised to be moderate and correlations with unrelated construct was hypothesised to be low. We also hypothesized that the change scores in physical function (ASPI/BASFI) would be able to distinguish between patients in the intervention and control group. Further, we hypothesized that the standard response mean (SRM) (defined in statistics) for change in physical function was at least moderate the intervention group. Responsiveness was considered acceptable if at least 75% of the hypotheses were confirmed.[33]

[H2]*Evaluation of Interpretability*

Interpretability is defined as the degree to which one can assign qualitative meaning to change in scores,[32] and was assessed according to COSMIN standard using the patient reported rating of perceived change in physical function as an anchor.[34] Patients were categorized into three groups according to their reported change; 1) very much better/much better, 2) a little better 3) no change/worse. The category "a little better" was considered as the minimal important change (MIC).

The ASAS20 response criterion is recommended to assess the effect of treatment in axSpA [35], and we therefore used this criterion to categorize responders. For ASPI, patients were classified as responders if they showed an improvement of ≥20% in at least one of the three subtests and no worsening ≥20% on the remaining subtests.[7] For BASFI, patients were classified as responders if they showed an improvement of ≥20% and ≥1 unit on BASFI score and no worsening of≥20% and ≥1 unit.[36,37]

[H2] Statistics

Descriptive statistics are provided as mean (SD) for continuous variables and numbers (percentages) for categorical variables. Differences between groups were examined with

independent sample t test or chi square test as appropriate. The scores of ASPI and BASFI were assessed for normality and missing values. For BASFI, floor and ceiling effect were defined as more than 15% of the patients scoring the lowest or highest score, respectively.[33] To test the hypotheses about correlations, Pearson correlation coefficients were computed. A correlation of $-0.3 \leq r \leq 0.3$ was considered low whereas a correlation of $r > 0.3$ or $r < -0.3$ was considered moderate.

To assess the ability to distinguish between patients in the intervention- and control-group the size of the area under the receiver operating characteristic (ROC) curve (AUC) was calculated. An AUC of at least 0.70 was considered adequate. SRM [mean change/SD change] in physical function for each group was calculated to provide information about the magnitude of the change. SRM was interpreted according to Cohens: small = 0.20, moderate = 0.50 and large = 0.80.

P-value < 0.05 was considered statistical significant. Statistical analyses were performed using SPSS version 21.0 (SPSS Inc, Chicago, Illinois, USA).

[H2]ROLE OF THE FUNDING SOURCE

The Norwegian Fund for Post-Graduate Training in Physiotherapy provided funding support. The funder played no role in the design, conduct, or reporting of this study.

[H1]RESULTS

A total of 67 patients underwent the assessment with ASPI, and of these 58 (87%) had complete data for ASPI and BASFI and are included in the analyses (Fig. 1). Excluded patients (n = 9) were statistically significant older mean age (SD) 53 (6.9) years, (p = 0.04) and a larger

proportion (44%) were smokers (p = 0.03). No other baseline characteristics and clinical features differ between the patients excluded and included in the analyses (data not shown).

Of the 58 included patients, 41% were male, mean (SD) age was 45 (10.7) years and 71% had radiological axial spondyloarthritis (Tab. 1). On average patients had high disease activity with a mean (SD) ASDAS of 2.6 (0.7) and 90% of the patients used medication for their axSpA. The most frequently patient-reported extra-spinal symptoms were peripheral joint pain and/or swelling (47%), enthesitis (38%) and uveitis (32%). There were no statistically significant differences between the intervention- and control group in baseline characteristics (Tab. 2).

The distribution of ASPI and BASFI scores at baseline and at 3 months are displayed in Figure 2. For BASFI, a score of <1 was present in 8 out of 58 (14%) at baseline.

[H2]*Responsiveness*

The predefined hypotheses regarding responsiveness of BASFI and ASPI are displayed in Table 2. Neither BASFI nor ASPI reached the predefined level of a confirmation of at least ≥75% of the hypotheses; for ASPI, 5 out of 8 (63%) (No. 1-2, 6-8) were confirmed, whereas for BASFI 2 out of 8 (25%) (No. 1, 8) were confirmed.

Change in ASPI and BASFI from baseline to 3 months with corresponding SMRs are shown in Table 3. In the intervention group, a larger SRM was found in ASPI (SRM = 1.08) than in BASFI (SRM = 0.90), whereas, in the control group, smaller SRM was found in ASPI (SMR = 0.06) than in BASFI (SRM = 0.27). The ability to discriminate between patients in the intervention and control group was AUC (95% CI) for ASPI 0.77 (0.64, 0.88) and for BASFI 0.69 (0.56, 0.83).

[H2]*Interpretability*

When patients reported their change in physical function at 3 months follow-up the distribution was the following; very much better/much better: 22 patients, a little better: 13 patients and no change/worse: 18 patients. In the intervention group, only one patient reported no change in physical function from baseline, while the remaining patients reported that their physical function was improved (Tab. 4). Change scores in ASPI and BASFI for these patients reported change groups are shown in Table 4. The category “a little better” was defined as the minimal important change, and we found this value to be 3.7 sec for ASPI and 0.8 point on a scale from 0 to 10 for BASFI.

The proportion of responders according to the ASAS20 response criteria for ASPI and BASFI are shown in Table 4. In the intervention group, 70% of the patients were categorized as responders with ASPI, whereas 43% were categorized as responders with BASFI, p=0.02. In the control group, 36% were categorized as responders with ASPI and 29% were categorized as responders with BASFI.

[H1] DISCUSSION

To the best of our knowledge, this is the first study to compare ASPI and BASFI in detecting change in physical function after a high intensity exercise intervention. The results indicate that ASPI is superior to BASFI in evaluating the effect of an exercise intervention on physical function in patients with axSpA. ASPI demonstrated better responsiveness for change in physical function and more patients in the exercise group was categorized as responders with ASPI than with the BASFI. In addition, a disadvantage with the BASFI was that we found a tendency towards a floor effect for this instrument.

Maintaining or improving physical function is regarded one of the main treatment targets in patients with axSpA,[35] and physical therapy with exercises is an important part of the

treatment. Hence, it is therefore of outmost importance to use measurement tools that can detect changes in physical function after an exercise intervention. Our finding that ASPI was superior to BASFI may be explained by the fact that SpA patients had more limited physical function at the time when the BASFI was developed. In line with the conclusion of a review of outcome measurement in SpA,[5] we observed a tendency towards a floor effect in BASFI. However, it has been requested that function in patients with axSpA should be measured within the dimensions of body structure, activity and participation according the International classification of function (ICF) model.[6] The BASFI includes items within the participation dimension, whereas in contrast, ASPI is a measure solely within the activity dimension. Hence, the BASFI and ASPI covers the dimensions of ICF differently, and an outcome measure within the activity dimension of ICF such as ASPI is probably most suitable for detecting changes in physical function after exercise interventions.

As expected, the hypothesis about the correlation between ASPI and BASFI was confirmed. These measurement instruments assess the same construct, as ASPI was developed based on BASFI and moderate correlations (r = 0.36-0.44) between single items in ASPI and BASFI have previously been shown in a cross-sectional study.[8] For ASPI, the hypothesis about moderate correlation between change scores was confirmed with regard to spinal mobility (BASMI), but not with cardiorespiratory fitness. This finding indicates that performing the ASPI-tasks (i.e. bending, putting on sock, getting up from the floor) are more strongly associated with changes in spine and hip mobility, than with cardiorespiratory fitness. In contrast, we found that BASFI was more correlated with constructs hypothesized to be unrelated, such as disease activity, back pain and fatigue than ASPI. These results are in line with a previous study, also showing that ASPI was more related with spinal mobility than with disease activity.[10] Further, previous studies have also shown that patient-reported

measures of physical function are more influenced by pain and exertion than performed based measures.[10,38] Hence, our findings indicate that BASFI to a larger extent than ASPI suffer from construct contamination, making BAFI less suited for detecting change in physical function.

We found that ASPI was superior to BASFI in distinguishing between patients in the intervention group and the control group, even if also the BASFI was close to proposed level of an adequate discrimination level.[33] Furthermore, the magnitude of change (SRM) was large in both ASPI (SRM; 1.08) and BASFI (SRM; 0.90) in the intervention group after high intensity exercise. In comparison, after treatment with biological medication for three months, ASPI showed a moderate magnitude of change (SRM for the different subtasks; 0.50 to 0.71).[7] However, it is reasonable to believe that 3 months of high intensity exercise will result in larger changes in physical function than biological medication.

The interpretability of ASPI has to our knowledge not been examined before. Our findings are consistent with findings in a recent study assessing the interpretability of BASFI, which showed that an improvement of 0.8 points was considered as “slightly improved” by the patients.[14] However, in the present study, the standard deviation was large in both instruments, indicating variance between patients, and the result should therefore be interpreted with caution.

The ASAS20 criteria is recommended for treatment evaluation in patients with axSpA, and according to these criteria, we found that 70% and 43% of the patients in the exercise group were responders in ASPI and BASFI, respectively. For ASPI, a similar proportion of responders (67%) has previously been reported after treatment with biological medication in patients with AS.[7,9] However, in contrast to our study, this study had a higher proportion of

responders also in BASFI (63%). The lower proportion of BASFI-responders in the current study may be explained by differences in the study populations, as less limitations in physical function were reported by the participants in the current study compared to the pharmacological-study (baseline mean BASFI 3.1 versus 5.4, respectively). In line with this, 14% of the patients had a score <1 at baseline, and further improvement in physical function cannot be detected according to the ASAS20 response criteria as this requires a change ≥1.[36,37] Hence, our finding suggests that a floor effect can occur when physical function is assessed with BASFI in populations with less impairment. This finding is in line with the result from a review showing that BASFI is less responsive in the relatively well patients.[5] Hence, based on the difference in the ability to detect responders between ASPI and BASFI after a physical therapy intervention, it can be calculated that the sample size of a clinical trial can be reduced by approximately 30% if ASPI is used as the primary outcome instead of BASFI. It should be noticed that since ASPI is a continuous measure, floor and ceiling effects was not tested. However, as the ASPI scores were almost normally distributed, there is no reason to believe that floor or ceiling effects affect ASPI.

Our study has several strengths. We complied with the COSMIN recommendation for conducting construct responsiveness in a longitudinal study, as a proportion of patients were expected to improve and the measurement instruments were not used as primary outcome. In addition, the responsiveness of the two instruments was measured at the same time in the same population. However, some limitations must be addressed. The sample size of this study was moderate (n = 58). Furthermore, defining hypotheses remains arbitrary regarding the number of predefined hypotheses and the magnitude and direction of the correlation coefficients defined. Therefore, we chose to use clearly predefined hypotheses as agreed by an expert group and included the same number of hypotheses for convergent and

discriminatory responsiveness. We applied the same hypotheses for both ASPI and BASFI even though one is a performance-based test and the other patient-reported measure. This can be questioned, as it is previously shown that BASFI have higher correlation with pain and fatigue than ASPI.[10] However, our aim was to evaluate the ability of the instruments to detect change in the construct to be measured (physical function), and we therefore applied the same hypotheses for both instruments. The choice of anchor for change may be debated, as we used scores of "a little better" along with "very much better and much better" as an indication of clinically important change. We argue, however, that it may be an important change if patients with a progressive and chronic disease, such as axSpA, perceive and report that their condition has become somewhat better, especially as their evaluation was given to an assessor blinded for the intervention.

In conclusion, our findings suggest that ASPI is preferable when evaluating physical function after exercise interventions in patients with axSpA. The performance-based test ASPI had higher construct responsiveness for physical function and was more sensitive to identify responders than the patient-reported measure BASFI after high intensity exercise.

Author Contributions and Acknowledgments:

Concept/idea/research design: C. Fongen, H. Dagfinrud, S. van Weely, K.B. Hagen, S.H. Sveaas

Writing: C. Fongen, H. Dagfinrud, A. Bilberg, S. van Weely, K.B. Hagen, S.H. Sveaas

Data collection: A. Bilberg, S.H. Sveaas

Data analysis: H. Dagfinrud, S. van Weely, K.B. Hagen, S.H. Sveaas

Project management: H. Dagfinrud, A. Bilberg, M.W. Johansen, K.B. Hagen, S.H. Sveaas

Providing participants: A. Bilberg, E. Pedersen, M.W. Johansen

Providing facilities/equipment: A. Bilberg, E. Pedersen, M.W. Johansen, K.B. Hagen

Providing institutional liaisons: K.B. Hagen

Consultation (including review of manuscript before submitting): C. Fongen, H. Dagfinrud, A. Bilberg, M.W. Johansen, K.B. Hagen, S.H. Sveaas

The authors gratefully thank all participating patients, study personnel at the different rheumatology departments, and the expert group for their contribution and The Norwegian Fund for Post-Graduate Training in Physiotherapy for financial support.

Ethics Approval

The trial was approved by the Regional Committee for Medical and Health Ethics of South East Norway (2015/86) and the Regional Ethical Review Board Gothenburg in Sweden (032-16). All participants provided their written informed consent before participation.

Funding

This study was funded by The Norwegian Fund for Post-Graduate Training in Physiotherapy.

Clinical Trial Registration

This study was registered at ClinicalTrials.gov (NCT02356874).

Disclosures

The authors completed the ICMJE Form for Disclosure of Potential Conflicts of Interest and reported no conflicts of interest.

REFERNCES

1 Sieper J, Poddubnyy D. Axial spondyloarthritis. *Lancet.* 2017.

2 van der Heijde D, Ramiro S, Landewe R, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. *Ann Rheum Dis.* 2017.

3 Sieper J, Rudwaleit M, Baraliakos X, et al. The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. *Ann Rheum Dis.* 2009;68 Suppl 2:ii1-44.

4 Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. *J Rheumatol.* 1994;21:2281-2285.

5 Zochling J. Measures of symptoms and disease status in ankylosing spondylitis: Ankylosing Spondylitis Disease Activity Score (ASDAS), Ankylosing Spondylitis Quality of Life Scale (ASQoL), Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI), Bath Ankylosing Spondylitis Global Score (BAS-G), Bath Ankylosing Spondylitis Metrology Index (BASMI), Dougados Functional Index (DFI), and Health Assessment Questionnaire for the Spondylarthropathies (HAQ-S). *Arthritis Care Res (Hoboken).* 2011;63 Suppl 11:S47-58.

6 Braun J, Kiltz U, Baraliakos X, van der Heijde D. Optimisation of rheumatology assessments - the actual situation in axial spondyloarthritis including ankylosing spondylitis. *Clin Exp Rheumatol.* 2014;32:S-96-104.

7 van Weely SF, Dekker J, Steultjens MP, et al. Objective evaluation of physical functioning after tumor necrosis factor inhibitory therapy in patients with ankylosing spondylitis: a selection of 3 feasible performance-based tests. *J Rheumatol.* 2015;42:623-629.

8 van Weely SF, van Denderen CJ, van der Horst-Bruinsma IE, et al. Reproducibility of performance measures of physical function based on the BASFI, in ankylosing spondylitis. *Rheumatology (Oxford).* 2009;48:1254-1260.

9 van Weely SF, van Denderen JC, Steultjens MP, et al. What do we miss? ASAS non-responders on anti-TNF therapy show improvement in performance-based physical function. *Rheumatology (Oxford).* 2013;52:1884-1889.

10 van Weely SF, van Denderen JC, Steultjens MP, et al. Moving instead of asking? Performance-based tests and BASFI-questionnaire measure different aspects of physical function in ankylosing spondylitis. *Arthritis Res Ther.* 2012;14:R52.

11 Sveaas SH, Bilberg A, Berg IJ, et al. High intensity exercise for 3 months reduces disease activity in axial spondyloarthritis (axSpA): a multicentre r trial of 100 patients. *Br J Sports Med.* 2019.

12 Rudwaleit M, van der Heijde D, Landewe R, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. *Ann Rheum Dis.* 2009;68:777-783.

13 Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc.* 2007;39:1423-1434.

14 van Tubergen A, Black PM, Coteur G. Are patient-reported outcome instruments for ankylosing spondylitis fit for purpose for the axial spondyloarthritis patient? A qualitative and psychometric analysis. *Rheumatology (Oxford).* 2015;54:1842-1851.

15 Jenkinson TR, Mallorie PA, Whitelock HC, Kennedy LG, Garrett SL, Calin A. Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index. *J Rheumatol.* 1994;21:1694-1698.

16 van der Heijde D, Deodhar A, Inman RD, Braun J, Hsu B, Mack M. Comparison of three methods for calculating the Bath Ankylosing Spondylitis Metrology Index in a randomized placebo-controlled study. *Arthritis Care Res (Hoboken).* 2012;64(12):1919-1922.

17 Gladman DD, Inman RD, Cook RJ, et al. International spondyloarthritis interobserver reliability exercise--the INSPIRE study: I. Assessment of spinal measures. *J Rheumatol.* 2007;34:1733-1739.

18 Haywood KL, Garratt AM, Jordan K, Dziedzic K, Dawes PT. Spinal mobility in ankylosing spondylitis: reliability, validity and responsiveness. *Rheumatology (Oxford).* 2004;43:750-757.

19 Wanders A, Landewe R, Dougados M, Mielants H, van der Linden S, van der Heijde D. Association between radiographic damage of the spine and spinal mobility for individual patients with ankylosing spondylitis: can assessment of spinal mobility be a proxy for radiographic evaluation? *Ann Rheum Dis.* 2005;64:988-994.

20 Machado P, Landewe R, Braun J, Hermann KG, Baker D, van der Heijde D. Both structural damage and inflammation of the spine contribute to impairment of spinal mobility in patients with ankylosing spondylitis. *Ann Rheum Dis.* 2010;69:1465-1470.

21 Balke B, Ware RW. An experimental study of physical fitness of Air Force personnel. *U S Armed Forces Med J.* 1959;10:675-688.

22 Sveaas SH, Berg IJ, Provan SA, et al. Efficacy of high intensity exercise on disease activity and cardiovascular risk in active axial spondyloarthritis: a randomized controlled pilot study. *PLoS One.* 2014;9:e108688.

23 Halvorsen S, Vollestad NK, Provan SA, et al. Cardiorespiratory fitness and cardiovascular risk in patients with ankylosing spondylitis: a cross-sectional comparative study. *Arthritis Care Res (Hoboken).* 2013;65:969-976.

24 Whaley MH BP, Otto RM, Armstron LE, . *ACSM's guidelines for exercise testing and prescription.* 7th ed. Philadelphia: Lippincott Williams & Wilkins 2006.

25 Bouchard C, Rankinen T. Individual differences in response to regular physical activity. *Med Sci Sports Exerc.* 2001;33:S446-451; discussion S452-443.

26 Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. *J Rheumatol.* 1994;21(:2286-2291.

27 Haywood KL, Garratt AM, Dawes PT. Patient-assessed health in ankylosing spondylitis: a structured review. *Rheumatology (Oxford).* 2005;44:577-586.

28 van der Heijde D, Lie E, Kvien TK, et al. ASDAS, a highly discriminatory ASAS-endorsed disease activity score in patients with ankylosing spondylitis. *Ann Rheum Dis.* 2009;68(12):1811-1818.

29 Lukas C, Landewe R, Sieper J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis.* 2009;68:18-24.

30 Machado PM, Landewe R, Heijde DV. Ankylosing Spondylitis Disease Activity Score (ASDAS): 2018 update of the nomenclature for disease activity states. *Ann Rheum Dis.* 2018.

31 van Tubergen A, Coenen J, Landewe R, et al. Assessment of fatigue in patients with ankylosing spondylitis: a psychometric analysis. *Arthritis Rheum.* 2002;47:8-16.

32 Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63:737-745.

33 de Vet H, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine.* United Kingdom: Cambridge university press; 2011.

34 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10:407-415.

35 Smolen JS, Braun J, Dougados M, et al. Treating spondyloarthritis, including ankylosing spondylitis and psoriatic arthritis, to target: recommendations of an international task force. *Ann.Rheum.Dis.* 2013;73:6-16.

36 Anderson JJ, Baron G, van der Heijde D, Felson DT, Dougados M. Ankylosing spondylitis assessment group preliminary definition of short-term improvement in ankylosing spondylitis. *Arthritis Rheum.* 2001;44:1876-1886.

37 Brandt J, Listing J, Sieper J, Rudwaleit M, van der Heijde D, Braun J. Development and preselection of criteria for short term improvement after anti-TNF alpha treatment in ankylosing spondylitis. *Ann Rheum Dis.* 2004;63:1438-1444.

38 Terwee CB, van der Slikke RM, van Lummel RC, Benink RJ, Meijers WG, de Vet HC. Self-reported physical functioning was more influenced by pain than performance-based physical functioning in knee-osteoarthritis patients. *J Clin Epidemiol.* 2006;59:724-731.

Table 1.
Baseline Characteristics and Clinical Features of the 58 Patients With Axial Spondyloarthritis[a]

Characteristics	**N**	**n (%)**
Male	58	24 (41.4%)
Age in years, mean (SD)	58	45.0 (10.7)
Married (or living with partner)	58	43 (74.1%)
Education level (University)	58	32 (55.2%)
Smoking status (present smoking)	58	5 (9.6%)
Body mass index, ≥25.0	58	37 (63.8%)
Ankylosing spondylitis	58	41 (70.7%)
Disease activity		
ASDAS, mean (SD)	58	2.6 (0.7)
CRP, mean (SD)	58	2.9 (4.4)
Medication	58	
No medication		6 (10.3%)
Monotherapy		
NSAID		23 (39.7%)
TNFi		7 (12.1%)
sDMARD		1 (1.7%)
Combination therapy		
NSAID and TNFi		14 (24.1%)
TNFi and sDMARD		2 (3.4%)
NSAID, TNFi and sDMARD		5 (8.6%)
Extra-spinal symptoms		
Peripheral joint pain and/or swelling	58	27 (46.6%)
Enthesitis	56	21 (37.5%)
Uveitis	57	18 (31.6%)
Inflammatory bowel disease	58	7 (12.1%)
Psoriasis	56	5 (8.9%)

[a]Values are number (percentage) unless indicated. BASDAI = Bath ankylosing spondylitis disease activity index; BASFI = Bath ankylosing spondylitis functional index; BASMI = Bath ankylosing spondylitis metrology index; CRP = C-reactive protein; DMARD = disease modifying anti-inflammatory drug; NSAID = non steroid anti-inflammatory drug; SpA = spondyloarthritis; TNFi = tumor necrosis factors inhibitors (biological medication).

Table 2.
Predefined Hypotheses to Assess the Responsiveness of ASPI and BASFI in Patients With Axial Spondyloarthritis[a]

	Hypothesis	Rationale	Confirmed	
			ASPI	BASFI
1.	The correlations between change scores in ASPI and change scores of BASFI is moderate (r>0.30)	Performance-based and patient-reported physical function is similar constructs and ASPI is developed based on BASFI [8,10].	Yes r = 0.41	Yes
2.	The correlations between change scores in physical function (ASPI/ BASFI) and change scores in spinal mobility (BASMI) is moderate (r>0.30)	Physical function and spinal mobility are related but are dissimilar constructs	Yes r = 0.40	No r = 0.26
3.	The correlations between change scores in physical function (ASPI/ BASFI) and change scores in cardiorespiratory fitness (VO_2 peak) is moderate and negative (r< -0.30)	Physical function and cardiorespiratory fitness are related but are dissimilar constructs	No r = -0.23	No r = -0.26
4.	The correlations between change scores in physical function (ASPI/ BASFI) and change scores in patient-reported disease activity (BASDAI) is low (r≤0.30)	Physical function and disease activity are unrelated constructs	No r = 0.34	No r = 0.57
5.	The correlations between change scores in physical function (ASPI/ BASFI) and change scores in back pain (NRS 0-10) is low (r≤0.30)	Physical function and back pain are unrelated constructs	No R = 0.31	No r = 0.45
6.	The correlations between change scores in physical function (ASPI/ BASFI) and change scores in fatigue (NRS 0-10) is low (r≤0.30)	Physical activity and fatigue are unrelated constructs	Yes r = 0.29	No r = 0.34
7.	Change scores in physical function (ASPI/ BASFI) are expected to distinguish between patients in the intervention and control group (AUC>0.70)	In order to be considered as responsive, the instrument should be able to discriminate between patients in the intervention and control group after a high intensity exercise intervention	Yes AUC = 0.77	No AUC = 0.69
8.	The SRM in physical function (ASPI/ BASFI) is at least moderate (≥ 0.50) in the intervention group	The SRM in ASPI after treatment with TNFi in patients with AS have been shown to be moderate [7]	Yes SRM = 1.08	Yes SRM = 0.84
		Number of accepted hypotheses, n (%)	5 (63%)	2 (25%)

[a]ASPI = ankylosing spondylitis performance-based improvement; AUC = area under curve; BASDAI = Bath ankylosing spondylitis disease activity index; BASFI = Bath ankylosing spondylitis functional index; BASMI = Bath ankylosing spondylitis metrology index; NRS = numeric rating scale; r = Pearson correlation coefficient; SRM = standard response mean [mean change/SD change]; TNFi = tumor necrosis factors inhibitors (biological medication)

Table 3.
Change in ASPI, BASFI, and Measurements Included in the Hypotheses During the Intervention Period in the Intervention and the Control Group[a]

Test	Baseline	3 Months	Change Scores	SRM
Performance-based physical function				
ASPI total (time in seconds, shorter time indicates better function)				
Intervention	33.1 (10.7)	26.6 (8.7)	-6.5 (6.0)	-1.08
Control	29.2 (6.8)	28.9 (8.9)	-0.3 (5.4)	-0.06
ASPI sub-test (time in seconds, shorter time indicates better physical function)				
BENDING, PICKING UP PENS FROM THE FLOOR				
Intervention	15.3 (4.0)	12.3 (2.7)	-3.0 (2.6)	-1.17
Control	15.1 (4.0)	14.4 (4.9)	-0.8 (3.7)	-0.21
PUTTING ON SOCKS				
Intervention	12.6 (5.8)	10.3 (5.4)	-2.3 (4.7)	-0.49
Control	9.3 (2.6)	10.3 (4.9)	1.0 (3.8)	0.25
GETTING UP FROM THE FLOOR				
Intervention	5.1 (2.3)	3.9 (1.7)	-1.2 (1.2)	-0.99
Control	4.8 (1.6)	4.3 (1.5)	-0.5 (0.8)	-0.65
Patient-reported physical function				
BASFI (0-10, where the highest score represent most limited physical function)				
Intervention	2.9 (2.0)	1.7 (1.5)	-1.2 (1.3)	-0.90
Control	3.4 (1.8)	3.1 (1.7)	-0.3 (1.1)	-0.27
Measures included in the hypotheses				
BASMI (0-10, where the highest score represent most limited spinal mobility)				
Intervention	2.9 (1.4)	2.4 (1.3)	-0.5 (0.8)	-0.60
Control	2.5 (1.2)	2.4 (1.3)	-0.1 (0.5)	-0.24
VO_2 peak (ml/kg/min)				
Intervention	36.9 (4.5)	39.3 (5.7)	2.6 (2.5)	1.01
Control	35.5 (6.7)	35.2 (7.4)	0.3 (2.8)	0.11
BASDAI (0-10, where the highest score represent most disease activity)				
Intervention	4.8 (1.6)	3.2 (1.6)	-1.7 (1.9)	-0.89
Control	5.2 (1.4)	4.7 (1.4)	-0.6 (1.3)	-0.43
Back pain (NRS 0-10, where the highest score represent most back pain)				
Intervention	4.6 (2.0)	2.5 (2.0)	-2.0 (2.1)	-0.94
Control	5.0 (2.2)	4.3 (2.5)	-0.6 (2.1)	-0.30
Fatigue (NRS 0-10, where the highest score represent most fatigue)				
Intervention	6.0 (1.8)	3.7 (2.2)	-2.4 (2.6)	-0.90
Control	6.1 (1.8)	5.4 (1.7)	-0.7 (2.1)	-0.33

[a]Values are mean (SD) or SRM. Number of patients included in the measurements; all patients n=58, control n=28, intervention n=30 (in VO_2 peak; all patients n=55, control n=26, intervention n=29).
ASPI = ankylosing spondylitis performance index; BASDAI = Bath ankylosing spondylitis disease activity index; BASFI = Bath ankylosing spondylitis functional index; BASMI = Bath ankylosing spondylitis metrology index; SRM = standard response mean (mean change/SD change).

Table 4.

Interpretability of ASPI and BASFI[a]

Parameter	Mean Change Score (SD) for[b]:		No. (%) of Participants in:	
	ASPI (s)	BASFI (0–10)	Intervention Group (n = 29)	Control Group (n = 24)
Participant-reported change				
Very much better/much better	−6.5 (6.9)	−1.4 (1.4)	19 (65.5)	3 (12.5)
A little better (MIC)	−3.7 (6.2)	−0.8 (1.2)	9 (31.0)	4 (16.7)
No change/worse	−0.1 (5.0)	−0.2 (1.0)	1 (3.4)	17 (70.8)
Responders according to ASAS20 criteria				
ASPI				
Responder			21 (70.0)	10 (35.7)
Nonresponder			9 (30.0)	18 (64.3)
BASFI				
Responder			13 (43.3)	8 (28.6)
Nonresponder			17 (56.7)	20 (71.4)
P[c]			.02	.45

[a]The change in ASPI and BASFI are shown for the different categories of patient reported change. Further, the number of responders in ASPI and BASFI according to the ASAS20 criteria in the intervention- and the control group are shown. For ASPI, responders were defined as ≥ 20% on at least one of the three subtests and no worsening ≥20% on the remaining subtests. In BASFI, responders were defined as ≥20% and ≥ 1 unit intra-individual change and no worsening of≥20% and ≥1 unit. ASPI = ankylosing spondylitis performance-based improvement; BASFI = Bath ankylosing spondylitis functional index; ASAS20 response criteria = The Assessment of Spondyloarthritis international Society (ASAS) 20% response criteria for improvement; MIC = minimal important change was defined as the category 'a little better'.

[b]There were 22, 13, and 18 participants in the "very much better/much better," "a little better (MIC)," and "no change/worse" groups, respectively.

[c]As determined by χ^2 tests between (non-) responder in ASPI and BASFI within groups.

Figure 1. Flow-chart of the participants

ASPI = ankylosing spondylitis performance index; BASFI = Bath Ankylosing Spondylitis Functional Index.

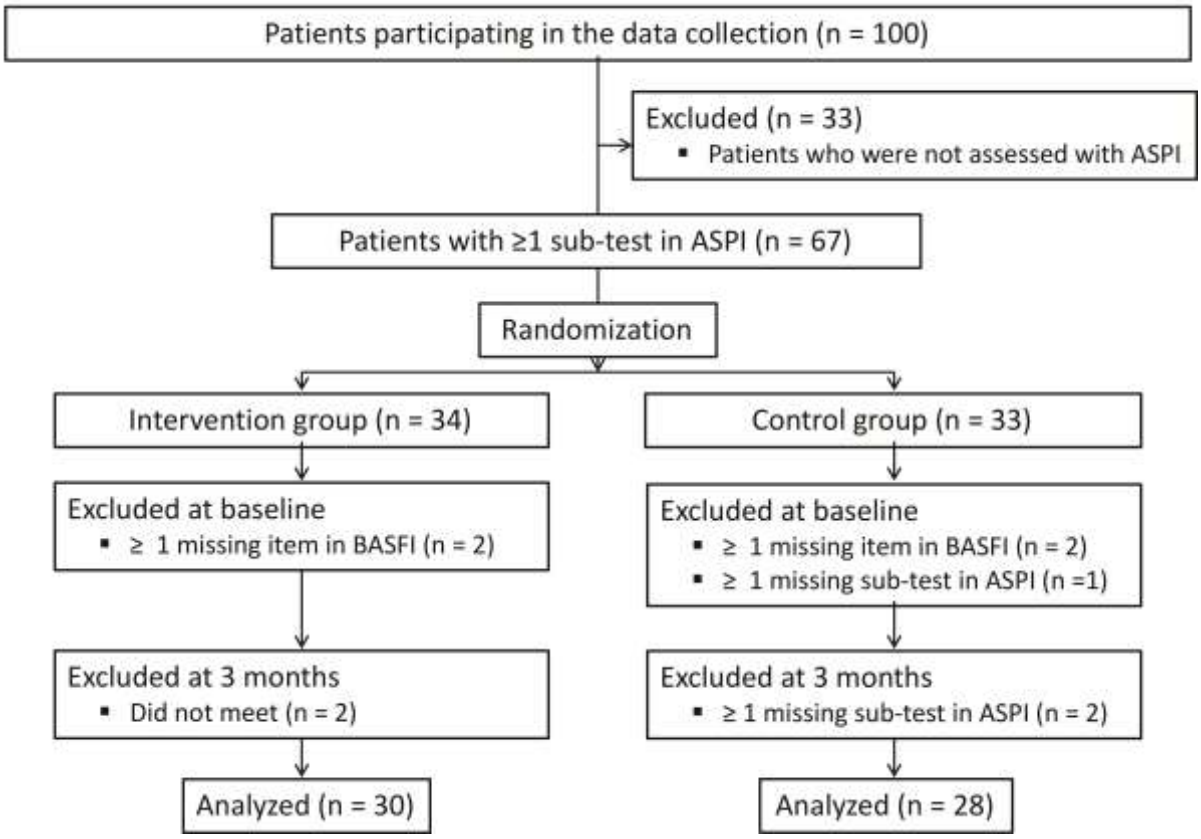

Figure 2. Distribution of BASFI scores and ASPI scores (time in seconds) at baseline and 3 months follow-up (n = 58). Lower values indicate better physical function in both measures.